Model bias estimation in data assimilation

Yannick TRÉMOLET

November 2005

---

1. Weak constraint 4D-Var
2. Results with systematic model error
3. Model bias and observation bias
4. Conclusions

4D Variational Data Assimilation

Variational data assimilation is based on the minimisation of:

$$
\begin{aligned}
J(x) &= \frac{1}{2}(x_0 - x_b)^T B^{-1} (x_0 - x_b) + \frac{1}{2}(y - \mathcal{H}(x))^T R^{-1} (y - \mathcal{H}(x)) \\
&+ \frac{1}{2}\mathcal{F}(x)^T C^{-1} \mathcal{F}(x)
\end{aligned}
$$

- x is the 4D state of the atmosphere over the assimilation window.
- $\mathcal{H}$ is a (sophisticated) 4D operator, accounting for the time dimension.
- $\mathcal{F}$ represents the remaining theoretical knowledge after background information has been accounted for. It may include equations governing the evolution of the atmosphere (model $\mathcal{M}$) and other constraints (DFI...).

Weak constraint 4D-Var

- Model error η is defined as: $\eta_i = x_i - \mathcal{M}_i(x_{i-1}) = \mathcal{F}_i(x)$,
- Choices of control vector $x = (x_i)_{i=0,\ldots,n}$ or $\chi = (x_0, (\eta_i)_{i=1,\ldots,n})$ are equivalent,
- The cost function becomes:

$$\begin{aligned} J(x) &= \frac{1}{2}(x_0 - x_b)^T B^{-1}(x_0 - x_b) \\ &+ \frac{1}{2}\sum_{i=0}^{n}(y_i - \mathcal{H}(x_i))^T R_i^{-1}(y_i - \mathcal{H}(x_i)) + \frac{1}{2}\sum_{i=1}^{n}\eta_i^T Q_i^{-1}\eta_i \end{aligned}$$

- Model error covariance matrix Q has to be defined.

Approximations

At the current operational resolution, the size of the problem is:

- Covariance matrix: 1.9×10^{16},
- 6×10^6 observations are available each day,
- At today's rate of observation it would take 9 million years to gather as many observations as there are parameters in Q.
- Weak constraint control variable: 194×10^6,
- Even if Q (or an approximation) was known, the number of degrees of freedom per analysis cycle is orders of magnitude larger than the number of observations.
- There is not enough information to solve the problem:

Approximations are required !!!

Choice of control variable

The 4D control variable $\{x_i\}_{i=0,\dots,n}$ can be replaced by:

x_0	x_0	x_0	x_0	x_0
x_i	η_i	$\eta_i = \eta$	$\eta_i = 0$	$\eta_i = 0$
$x_i \approx \mathcal{M}_i(x_{i-1})$	$x_i = \mathcal{M}_i(x_{i-1}) + \eta_i$	$x_i = \mathcal{M}_i(x_{i-1}) + \eta$	$x_i = \mathcal{M}_i(x_{i-1})$	$x_i = x_0$
$\Downarrow$	$\Downarrow$	$\Downarrow$	$\Downarrow$	$\Downarrow$
Weak constraint 4D-Var	Weak constraint 4D-Var	Practical Implementation	4D-Var	3D-Var

- Constant model error forcing makes weak constraint 4D-Var affordable,
- It is a representation of **systematic error**.

Model error covariance matrix

- The usual choice is $Q = \alpha B$.
- Linearisation in incremental formulation gives:

$$\delta x_n = M_n \dots M_1 \delta x_0 + \sum_{i=1}^{n} M_n \dots M_{i+1} \eta_i$$

- δx_0 can be identified with η_0.
- The solution of the analysis equation satisfies:

$$\delta x_0 = BH^T (R + HBH^T)^{-1} (y - \mathcal{H}(x_b))$$

$$\eta = QH^T (R + HQH^T)^{-1} (y - \mathcal{H}(x_b))$$

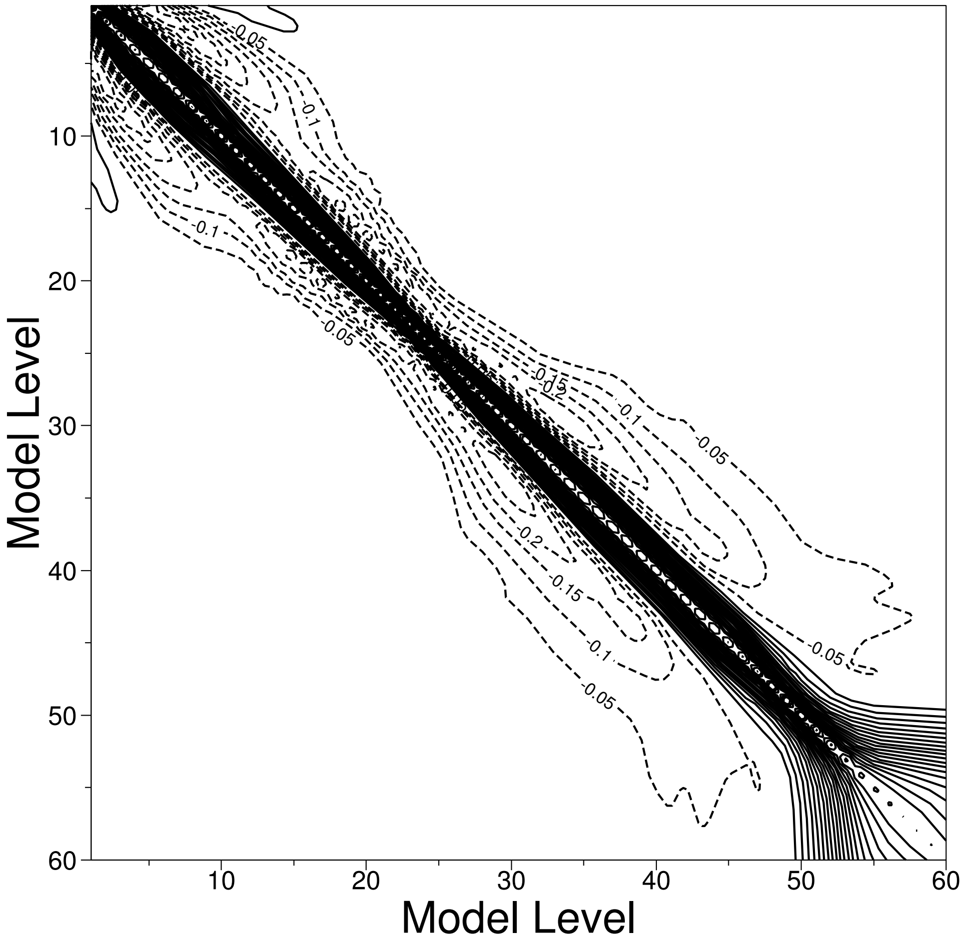
- If Q and B are proportional, δx_0 and η are constrained in the same directions, may be with different relative amplitudes.
- They both predominantly retrieve the same information: $Q = \alpha B$ is too limiting.

Generating a Model Error Covariance Matrix

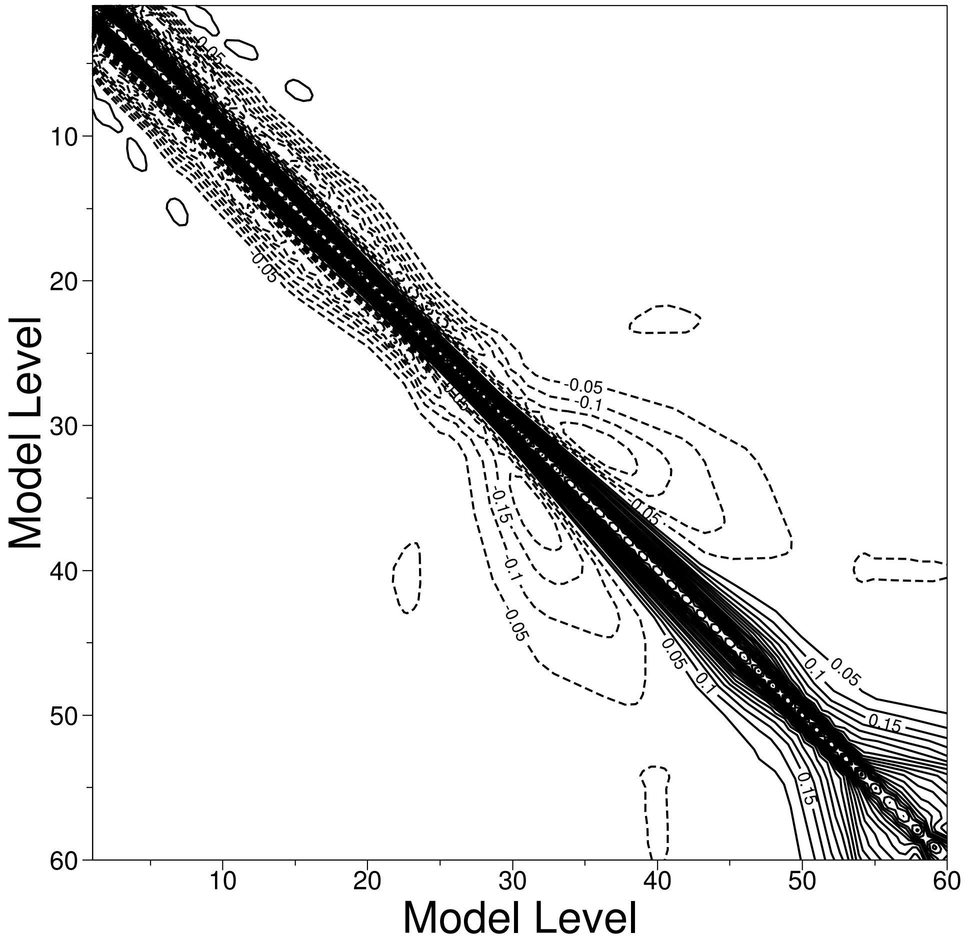
- B is estimated from an ensemble of 4D-Var assimilations.
- Considering the forecasts run from the 4D-Var members:
  - At a given step, each model state is supposed to represent the same *true* atmospheric state,
  - The tendencies from each of these model states should represent possible evolutions of the atmosphere from that same *true* atmospheric state,
  - The differences between these tendencies can be interpreted as possible uncertainties in the model or realisations of *model error*.
- Q can be estimated by applying the statistical model used for B to tendencies instead of analysis increments.

Average Temperature Vertical Correlations

Background Error

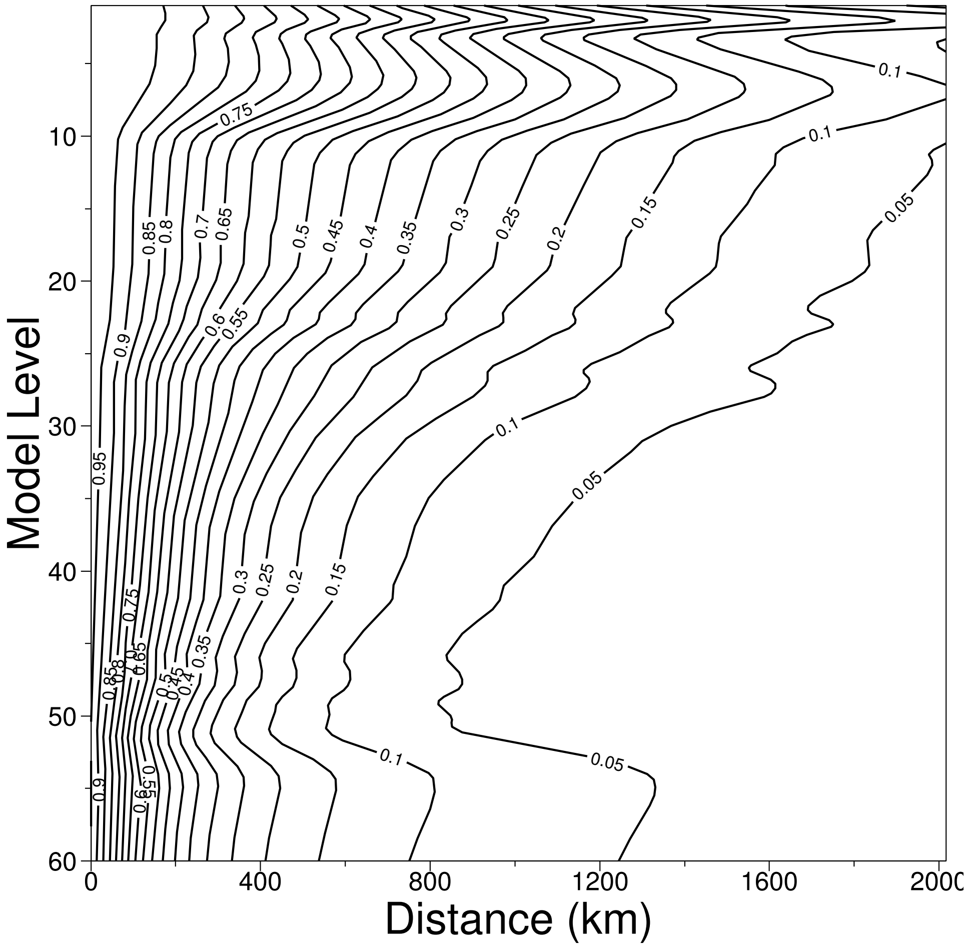


Model Error

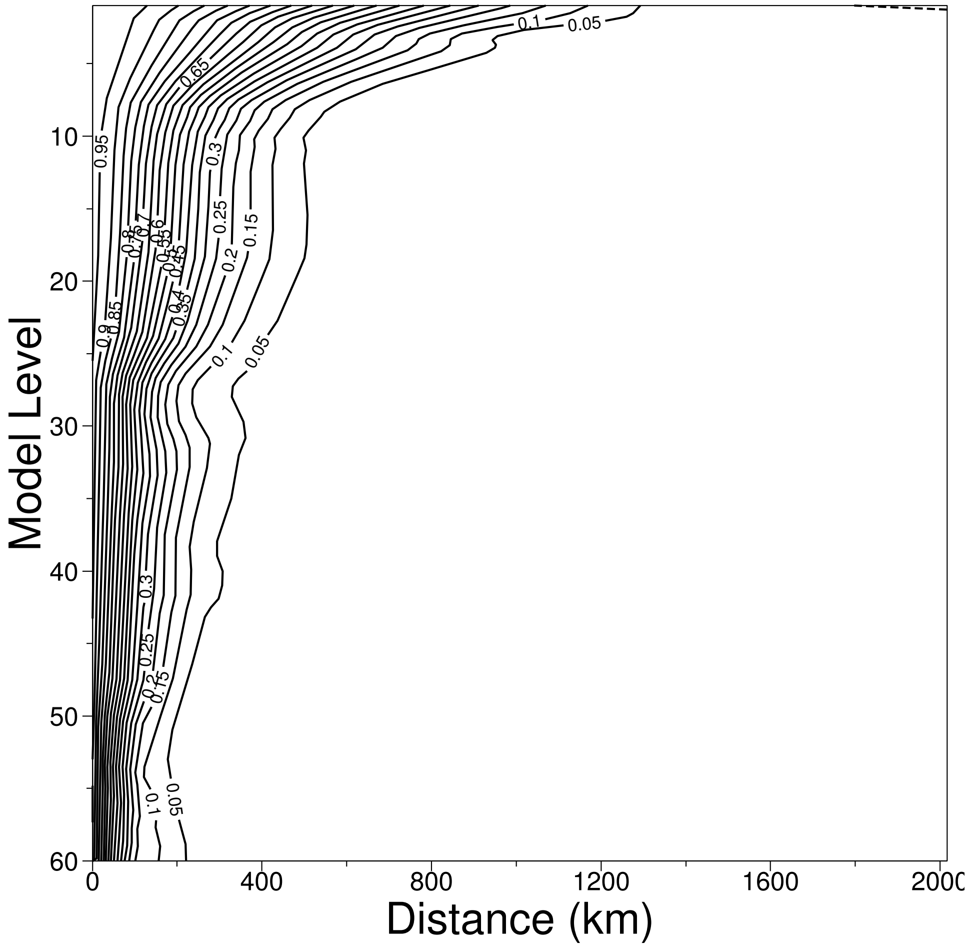


Temperature Horizontal Correlations

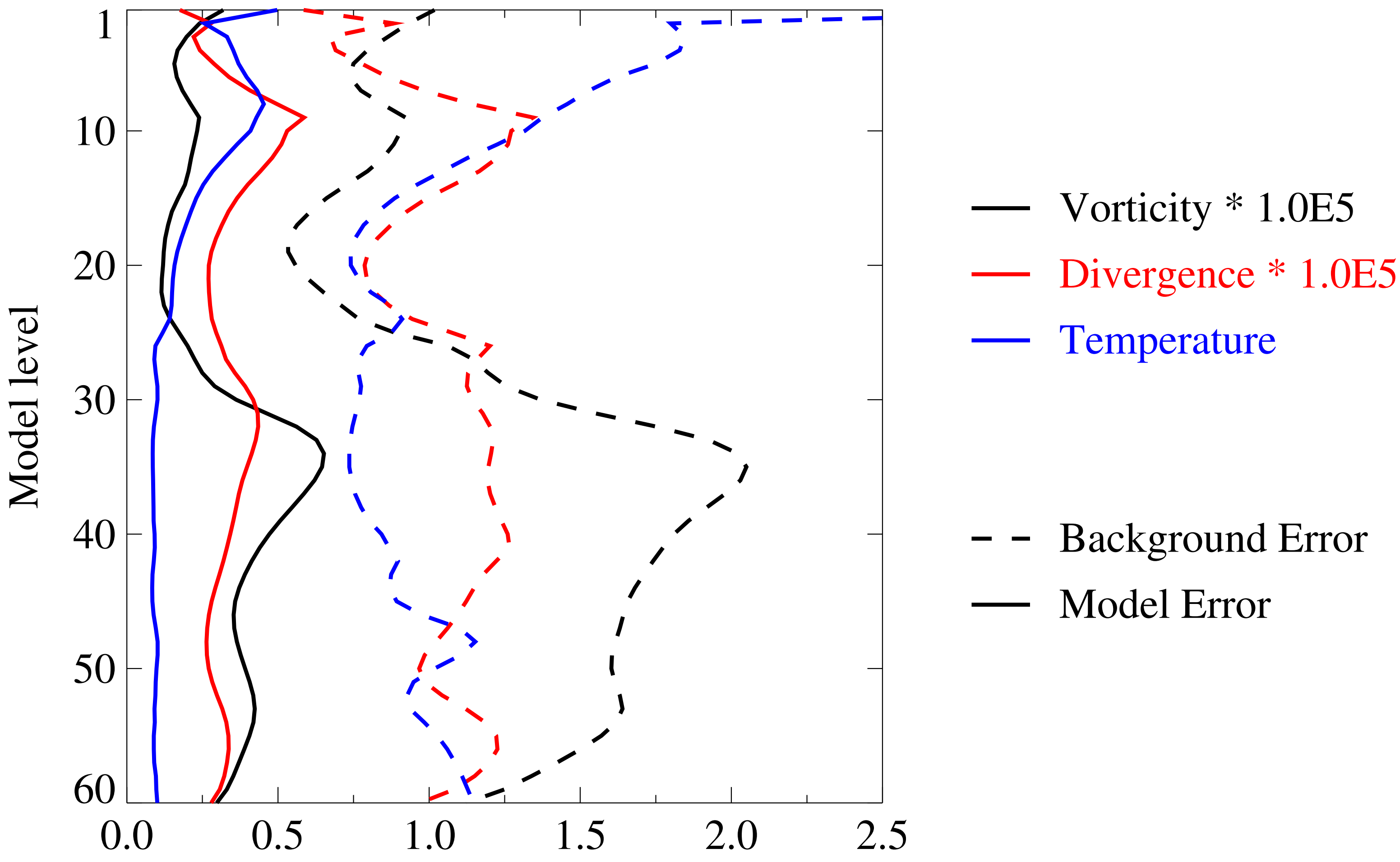
Background Error



Model Error



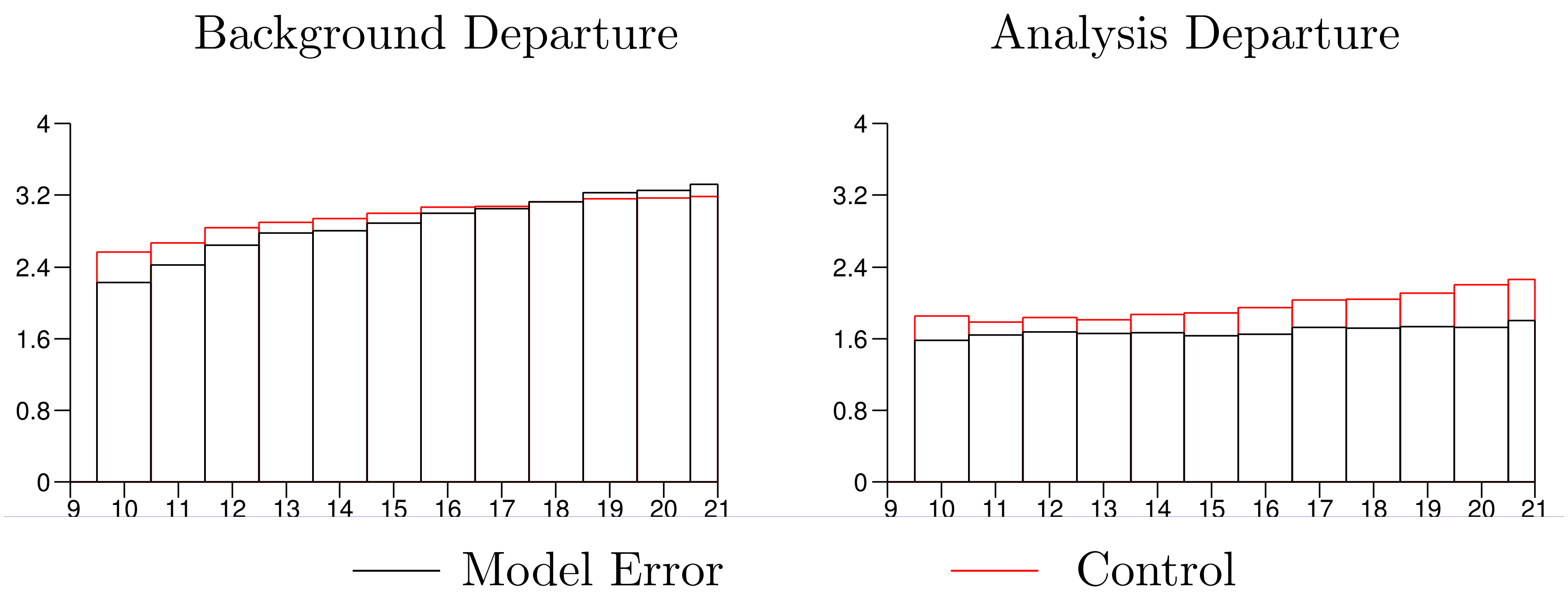
Average Standard Deviation Profiles



Q has narrower correlations and smaller amplitudes than B

Results: Fit to observations

`AMprofiler-windspeed Std Dev N.Amer`



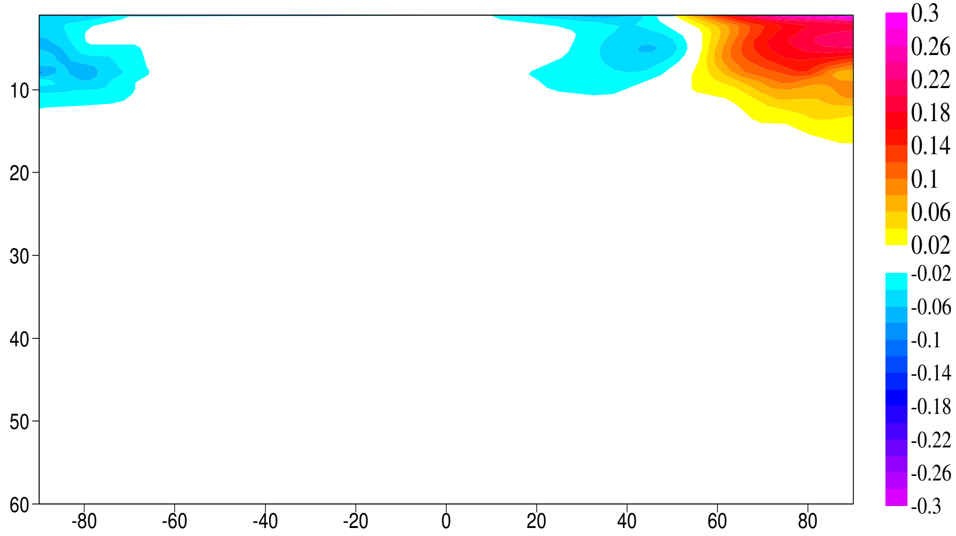
- Fit to observations is more uniform over the assimilation window.
- Background fit improved only at the start: error varies in time ?

Model Error Forcing

Zonal Mean Temperature

July 2004

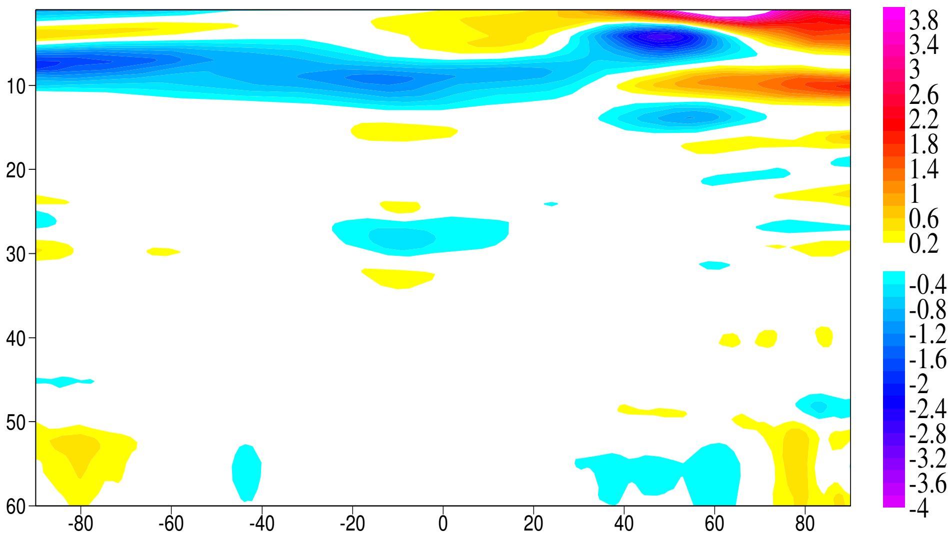
M.E. Forcing ⟶

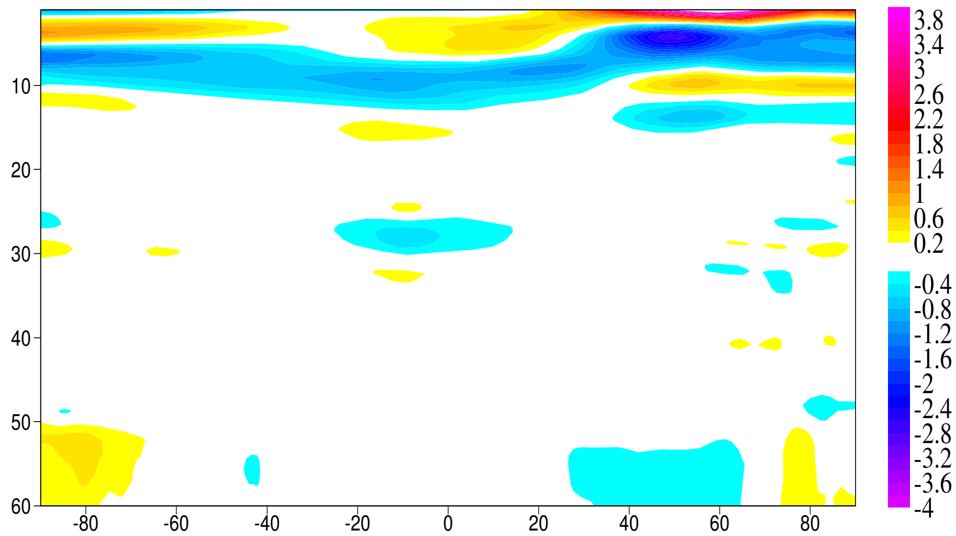


M.E. Mean Increment ↘

Control Mean Increment

↓





Mean Model Error Forcing

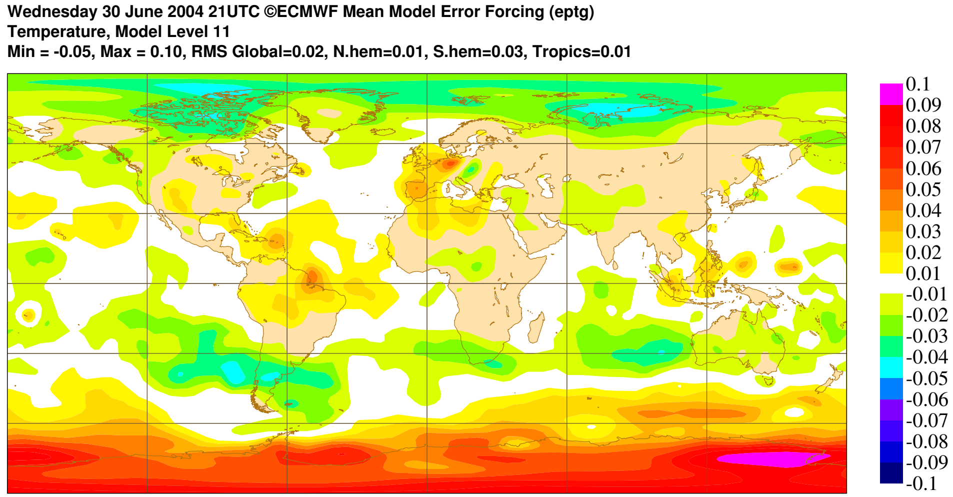
Model level 11 - Temperature

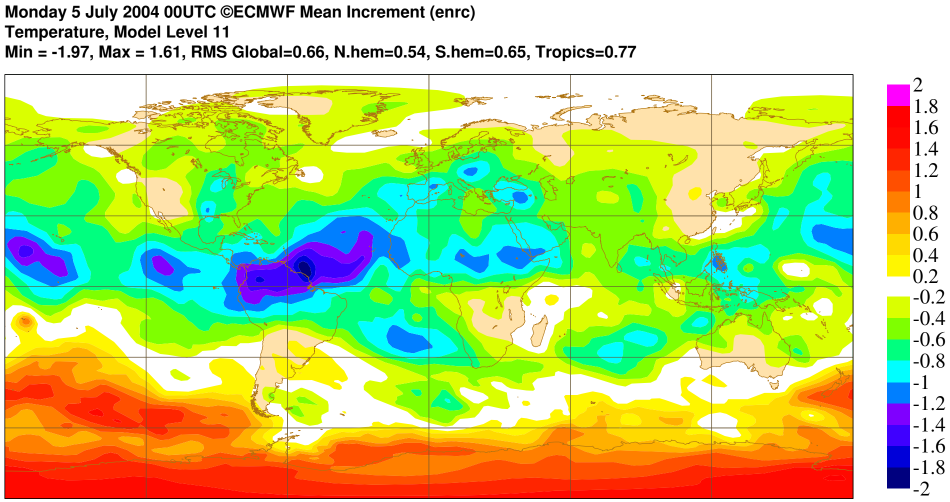
Mean M.E. Forcing ⟶

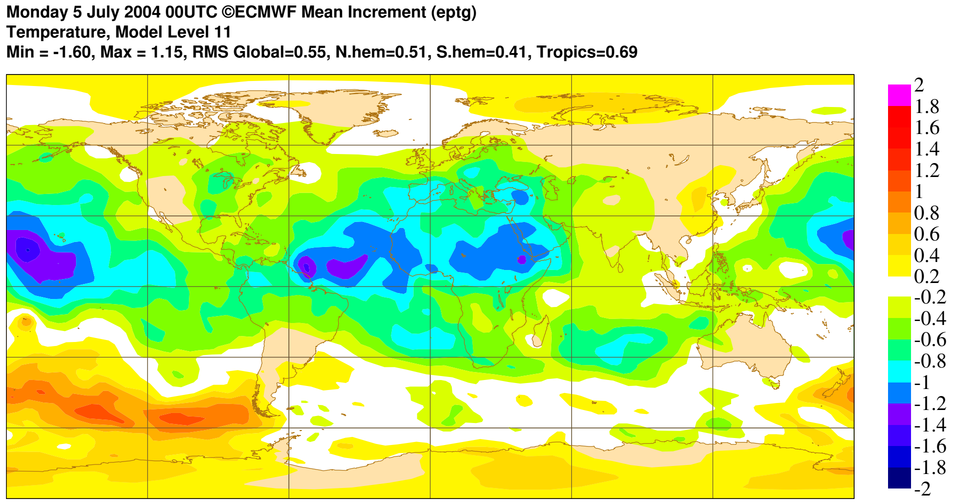
M.E. Mean Increment ↘

Control Mean Increment

↓

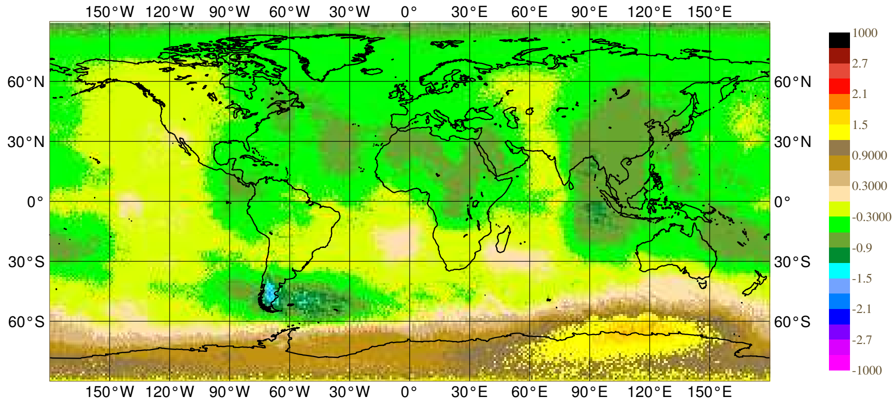




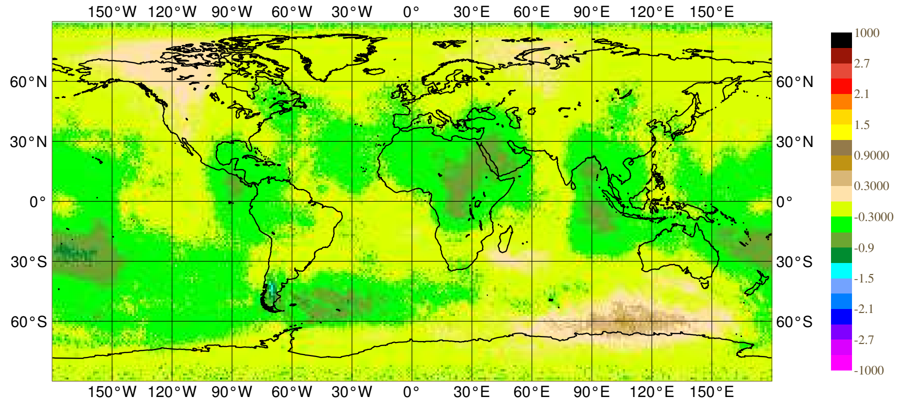


AMSU-A Departures

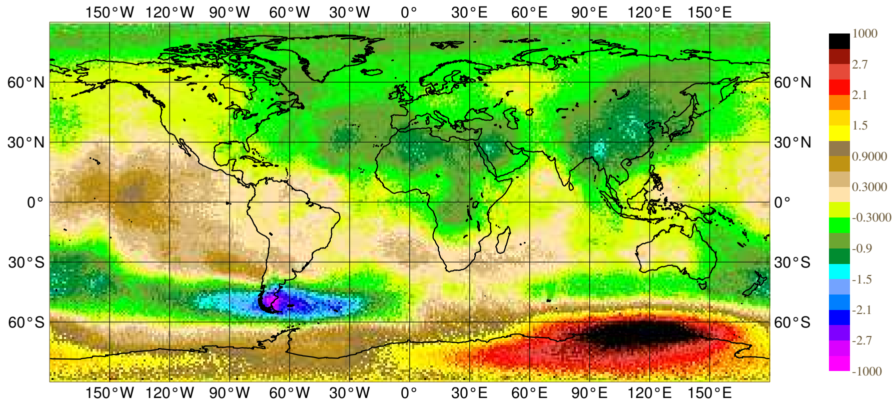
STATISTICS FOR RADIANCES FROM NOAA-16 / AMSU-A - 13
MEAN FIRST GUESS DEPARTURE (OBS-FG) (BCORR.) (CLEAR)
DATA PERIOD = 2004070200 - 2004073118 , HOUR = ALL
EXP = ENRC
Min: -1.9618 Max: 2.7 Mean: -0.169506



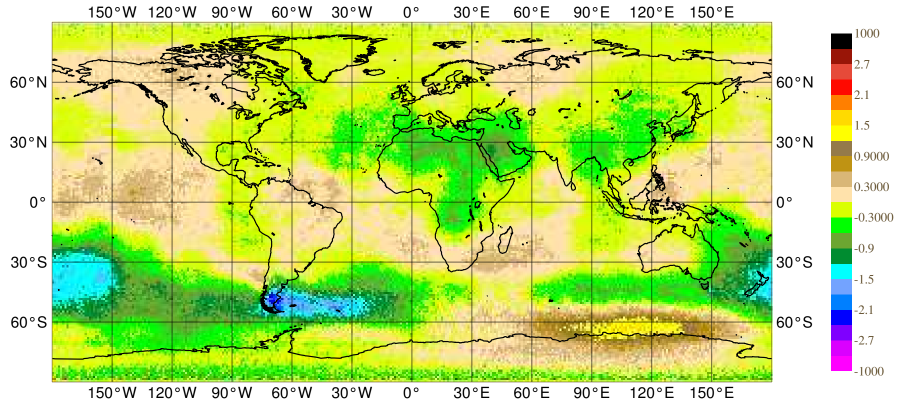
STATISTICS FOR RADIANCES FROM NOAA-16 / AMSU-A - 13
MEAN FIRST GUESS DEPARTURE (OBS-FG) (BCORR.) (CLEAR)
DATA PERIOD = 2004070100 - 2004073118 , HOUR = ALL
EXP = EPTG
Min: -1.6688 Max: 0.8 Mean: -0.231773



STATISTICS FOR RADIANCES FROM NOAA-16 / AMSU-A - 14
MEAN FIRST GUESS DEPARTURE (OBS-FG) (BCORR.) (CLEAR)
DATA PERIOD = 2004070200 - 2004073118 , HOUR = ALL
EXP = ENRC
Min: -3.3564 Max: 5.46 Mean: 0.006309

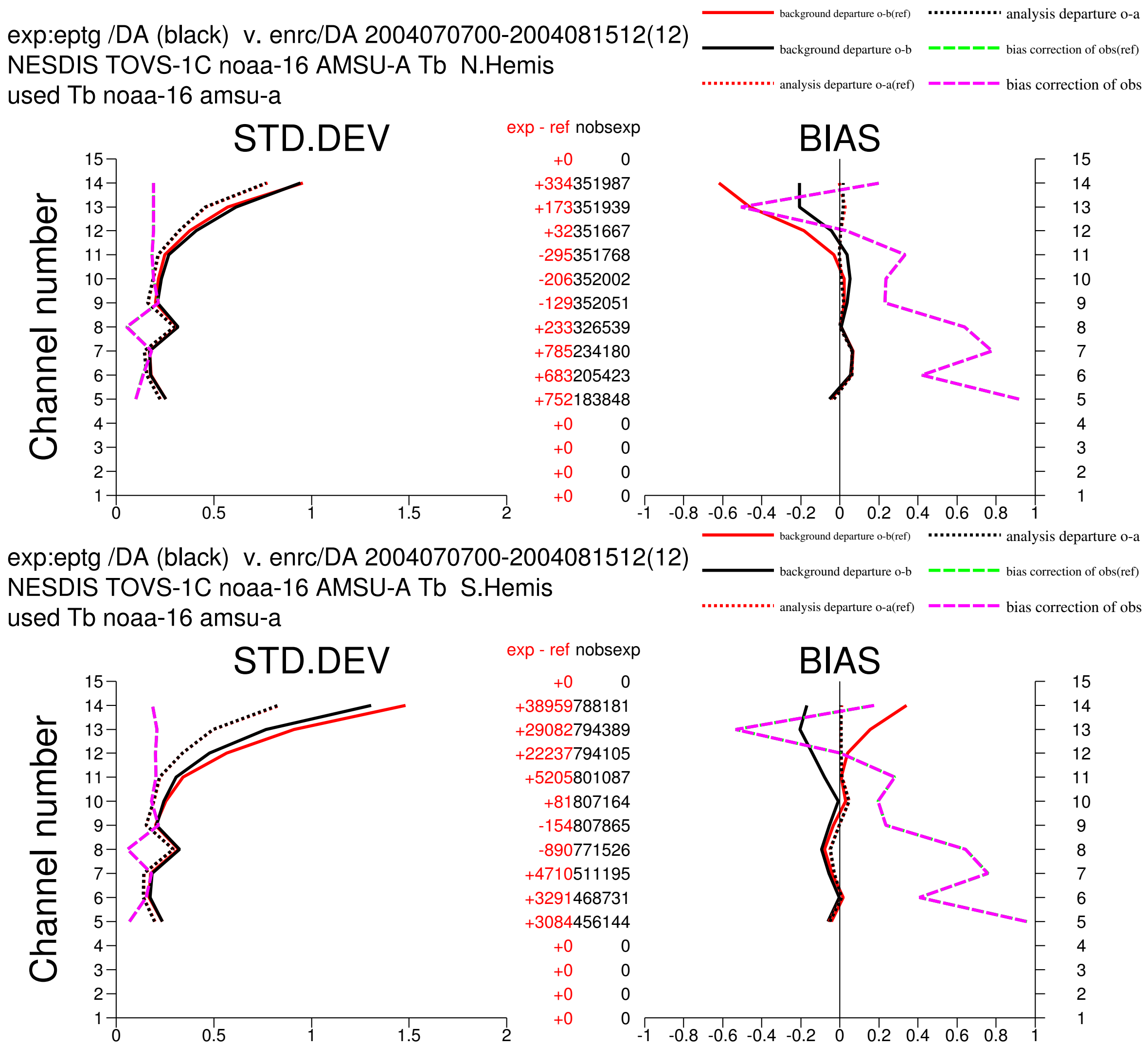


STATISTICS FOR RADIANCES FROM NOAA-16 / AMSU-A - 14
MEAN FIRST GUESS DEPARTURE (OBS-FG) (BCORR.) (CLEAR)
DATA PERIOD = 2004070100 - 2004073118 , HOUR = ALL
EXP = EPTG
Min: -2.6 Max: 2.16 Mean: -0.111883



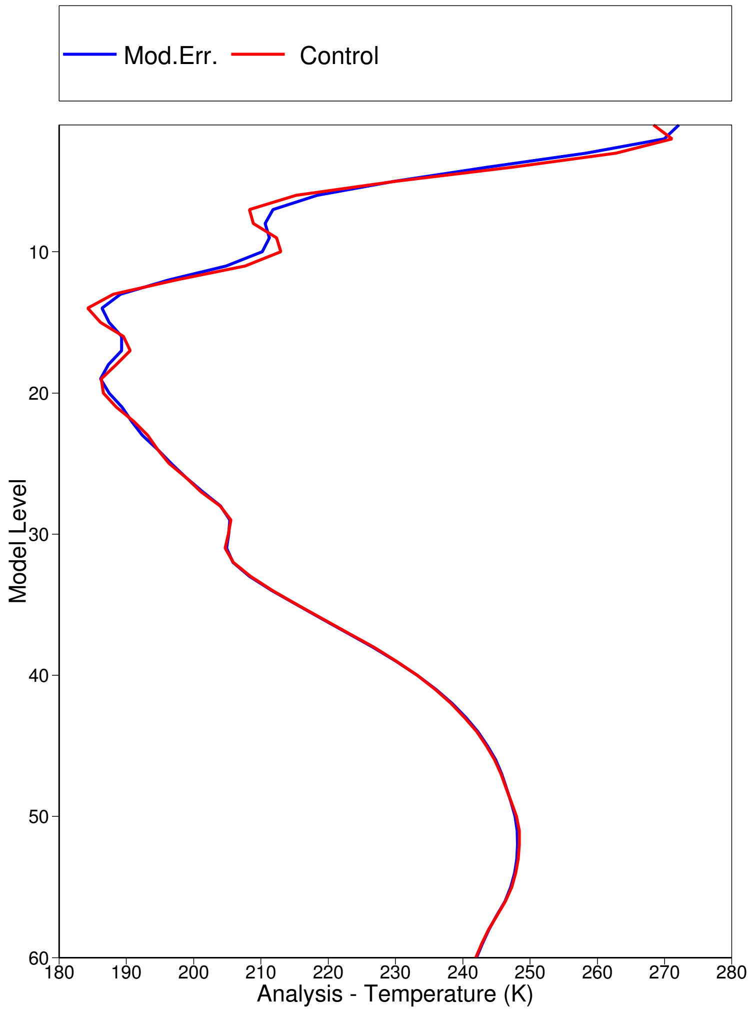
Control Model Error

AMSU-A Statistics

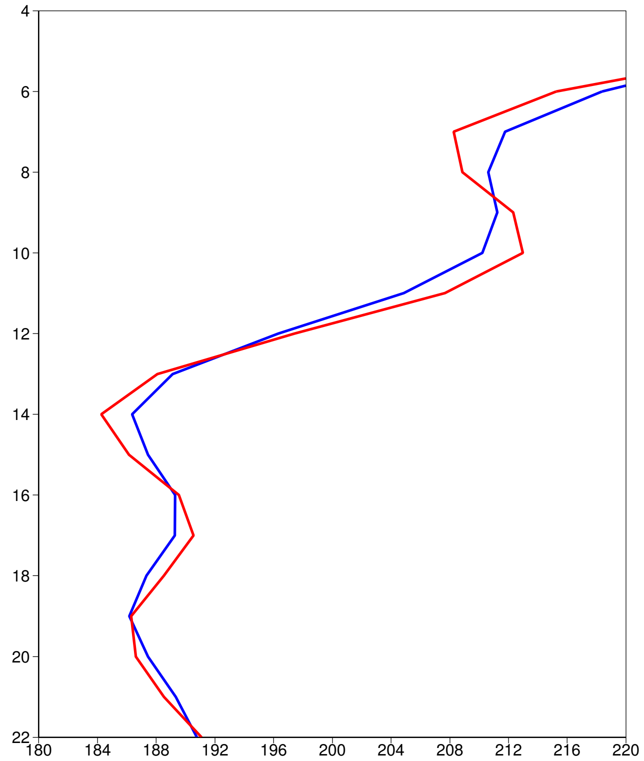


- Bias is more uniform,
- BG std. dev. is reduced in SH,
- More data is used.

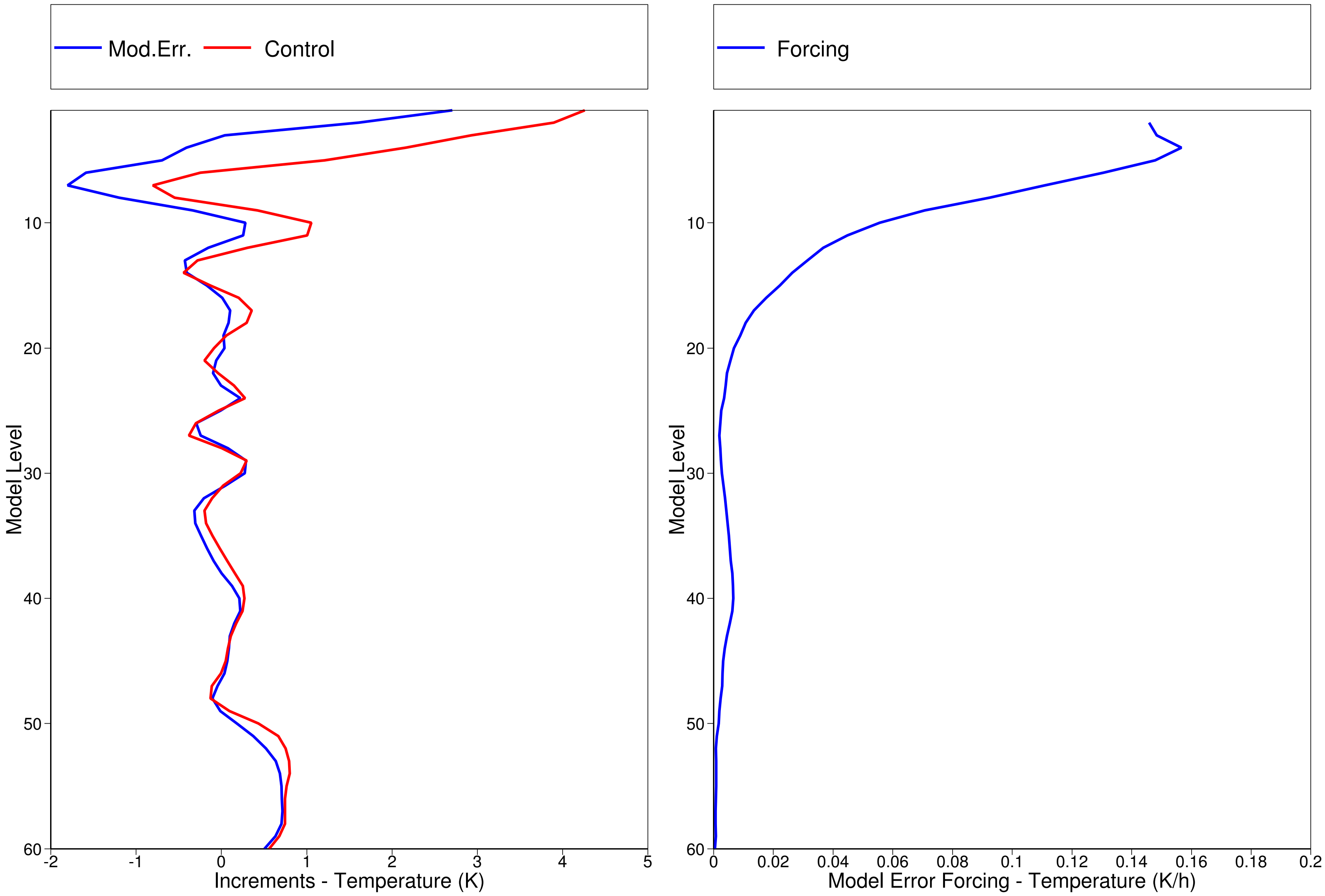
Temperature Analysis Profile



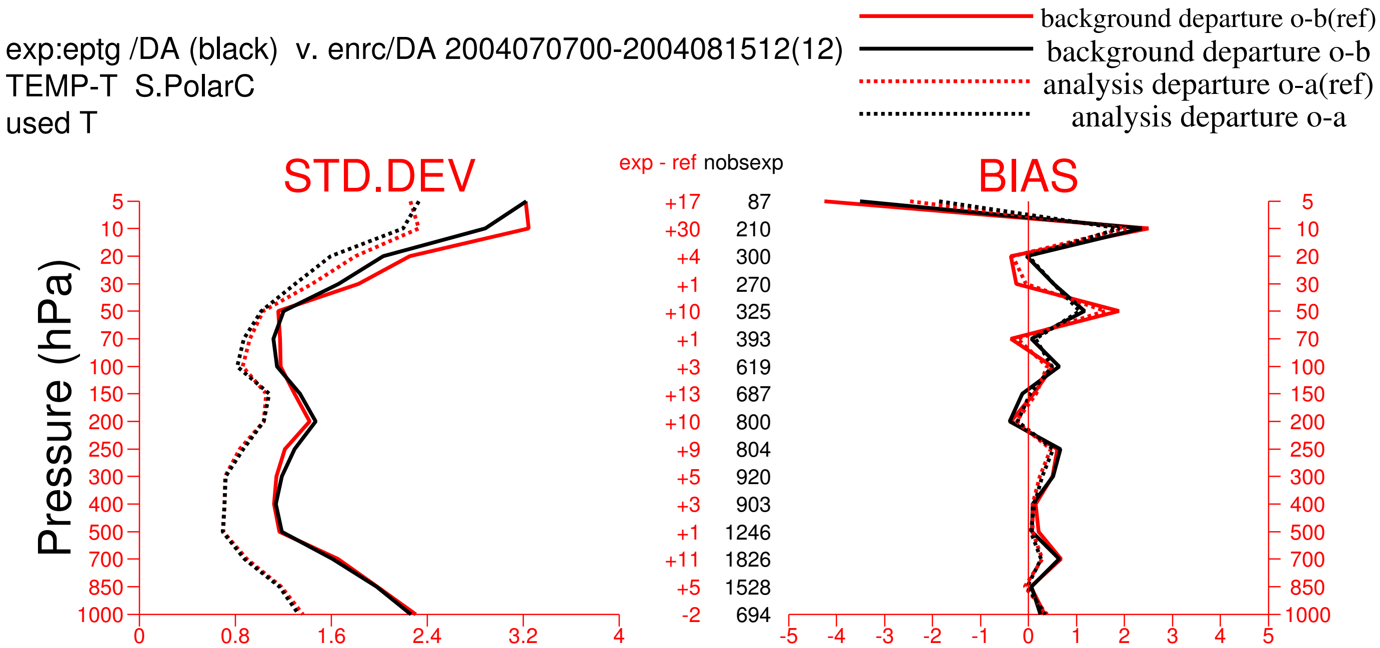
⟹



Temperature Increments Profile



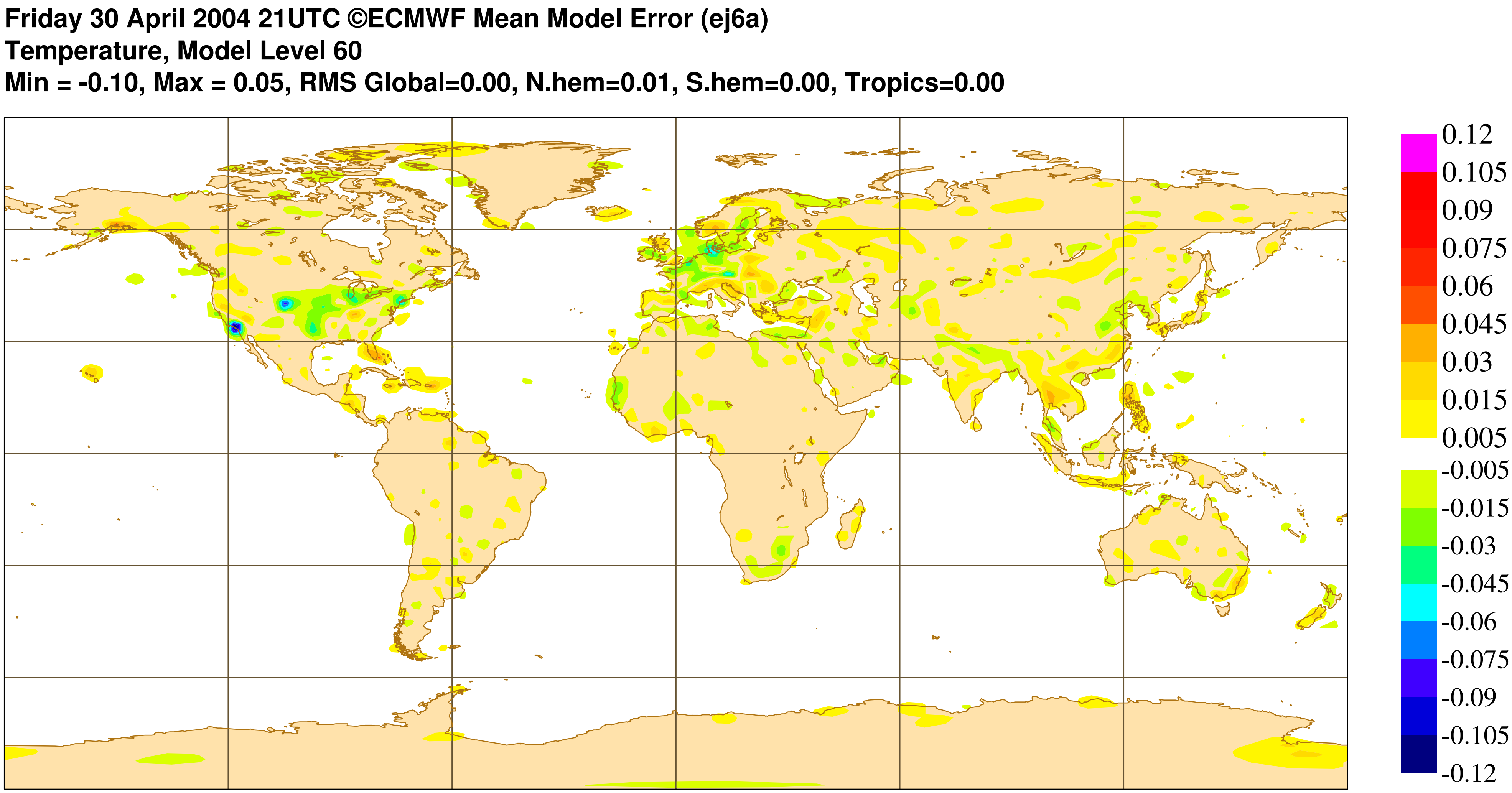
Fit to radiosonde data (cross-validation)



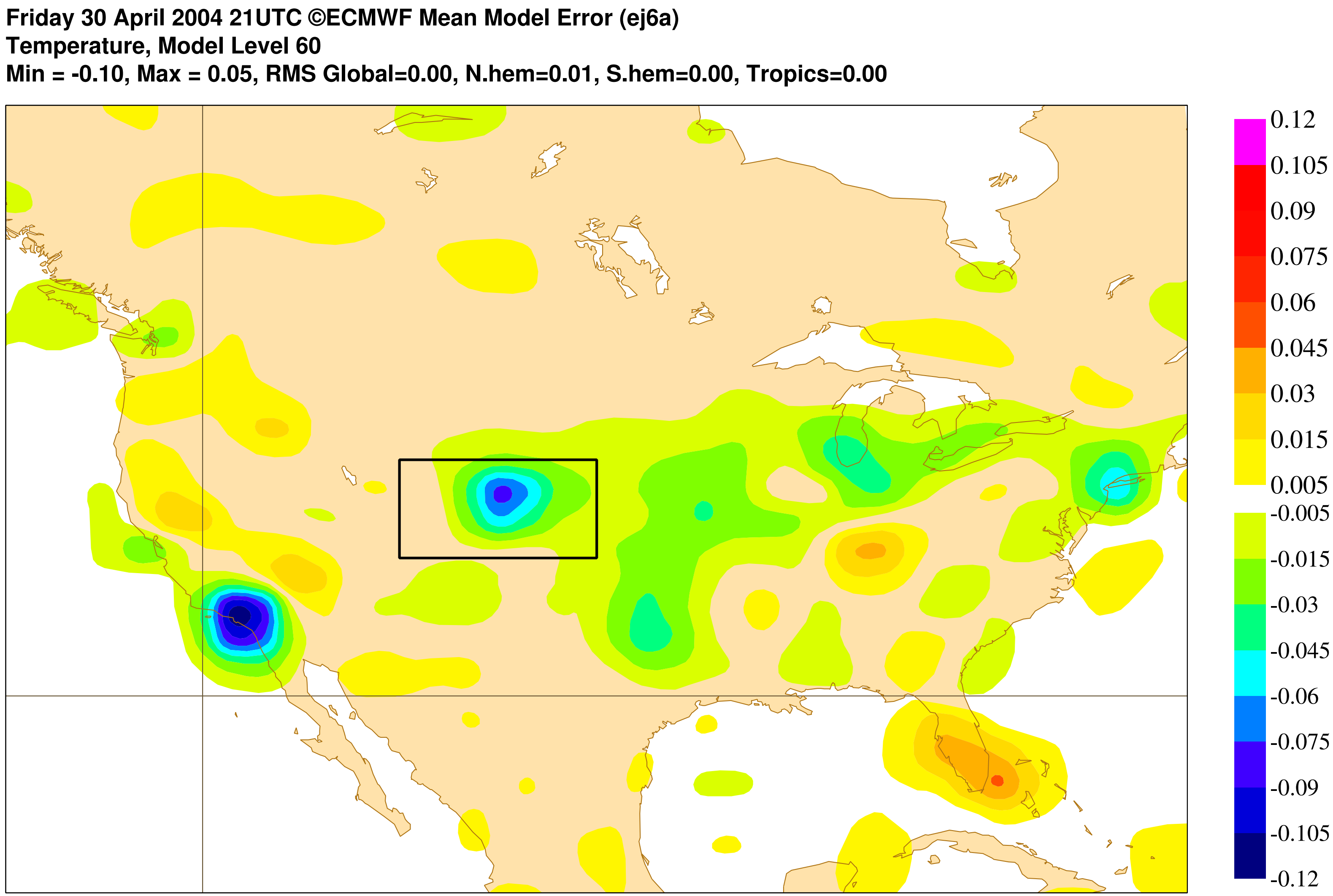
- Oscillations in bias are reduced,
- Std. deviation is reduced above 50 hPa (bg and an).

One source of model error was corrected by the forcing term.

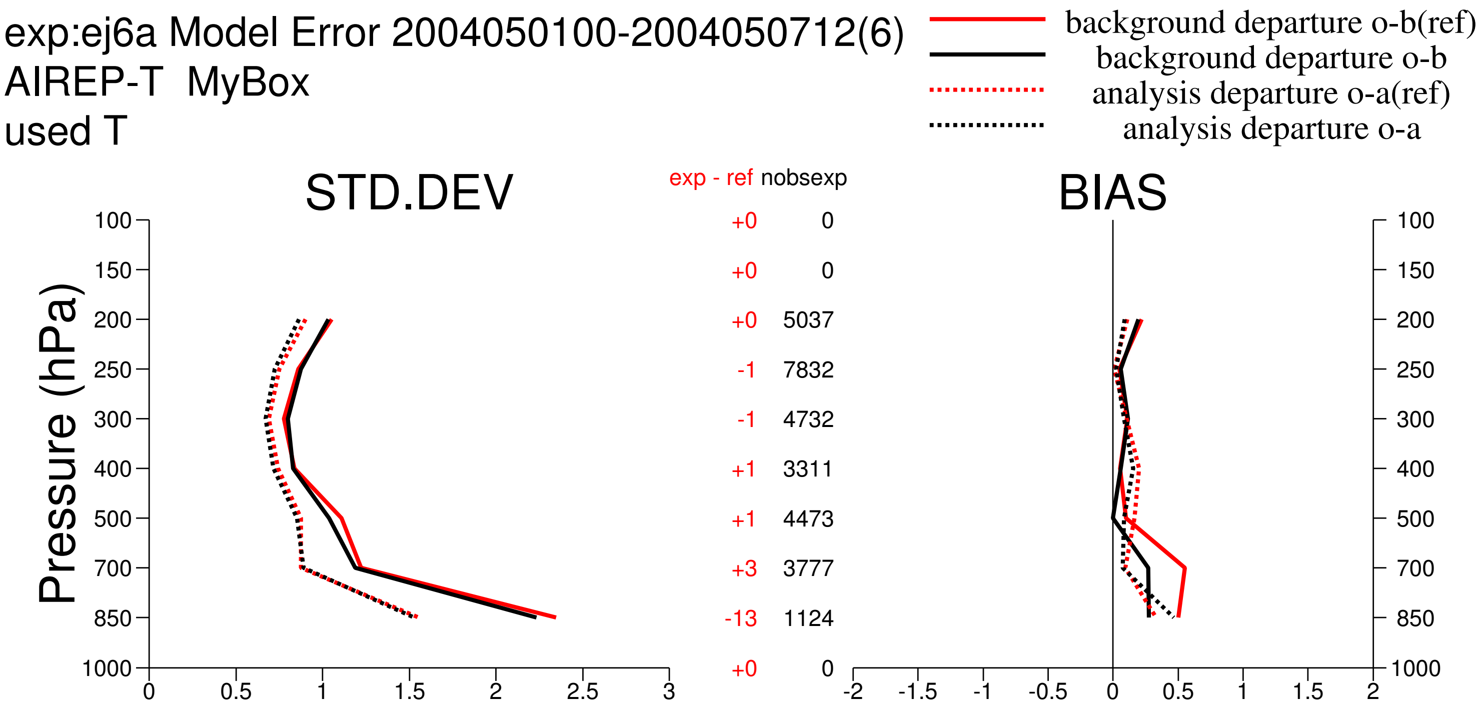
Low Level Mean Model Error Forcing



Low Level Mean Model Error Forcing

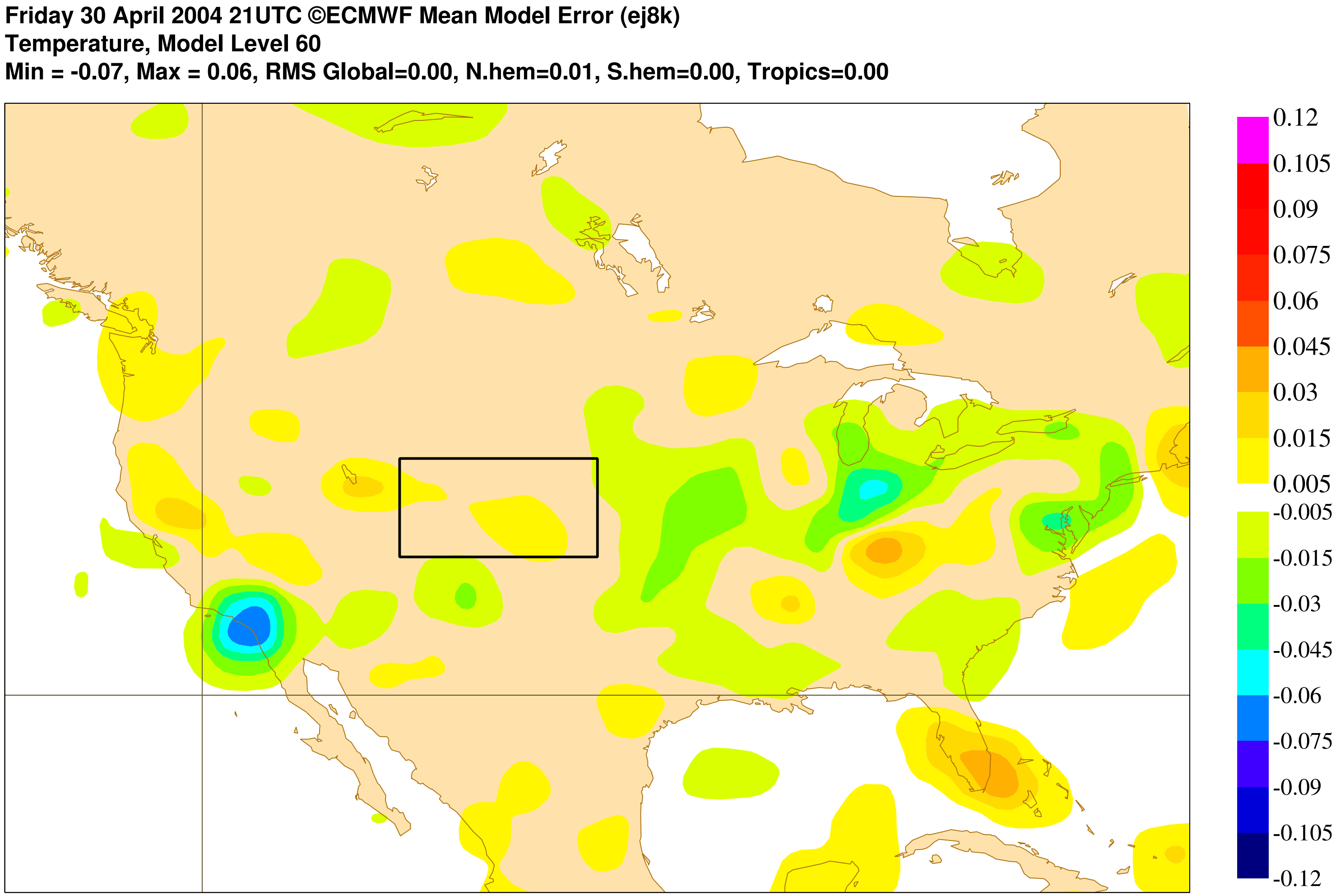


Fit to Observation with Model Error



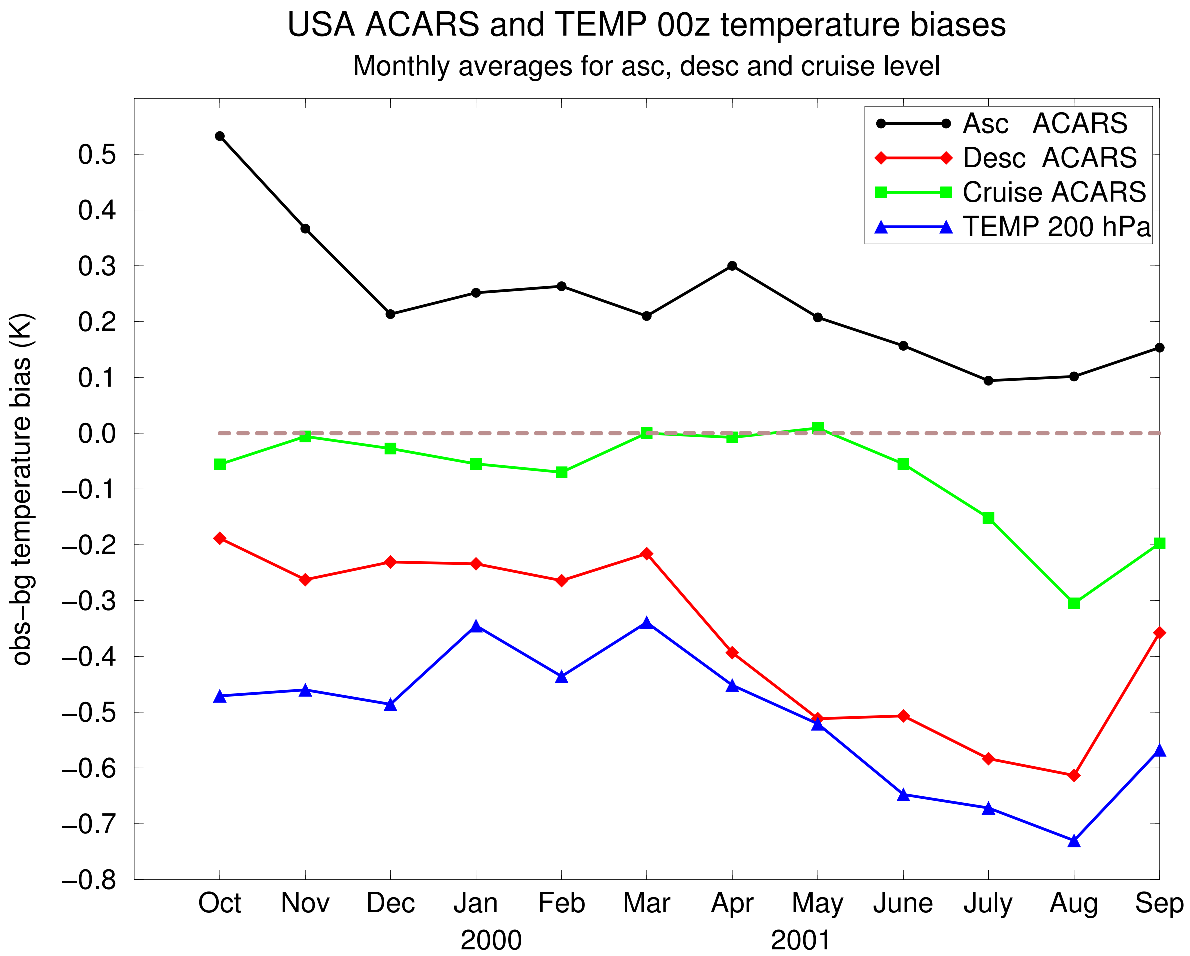
The bias for aircraft low level temperature observations was reduced.

Low Level Mean Model Error Forcing



Removing aircraft data in the box eliminates the (spurious?) forcing.

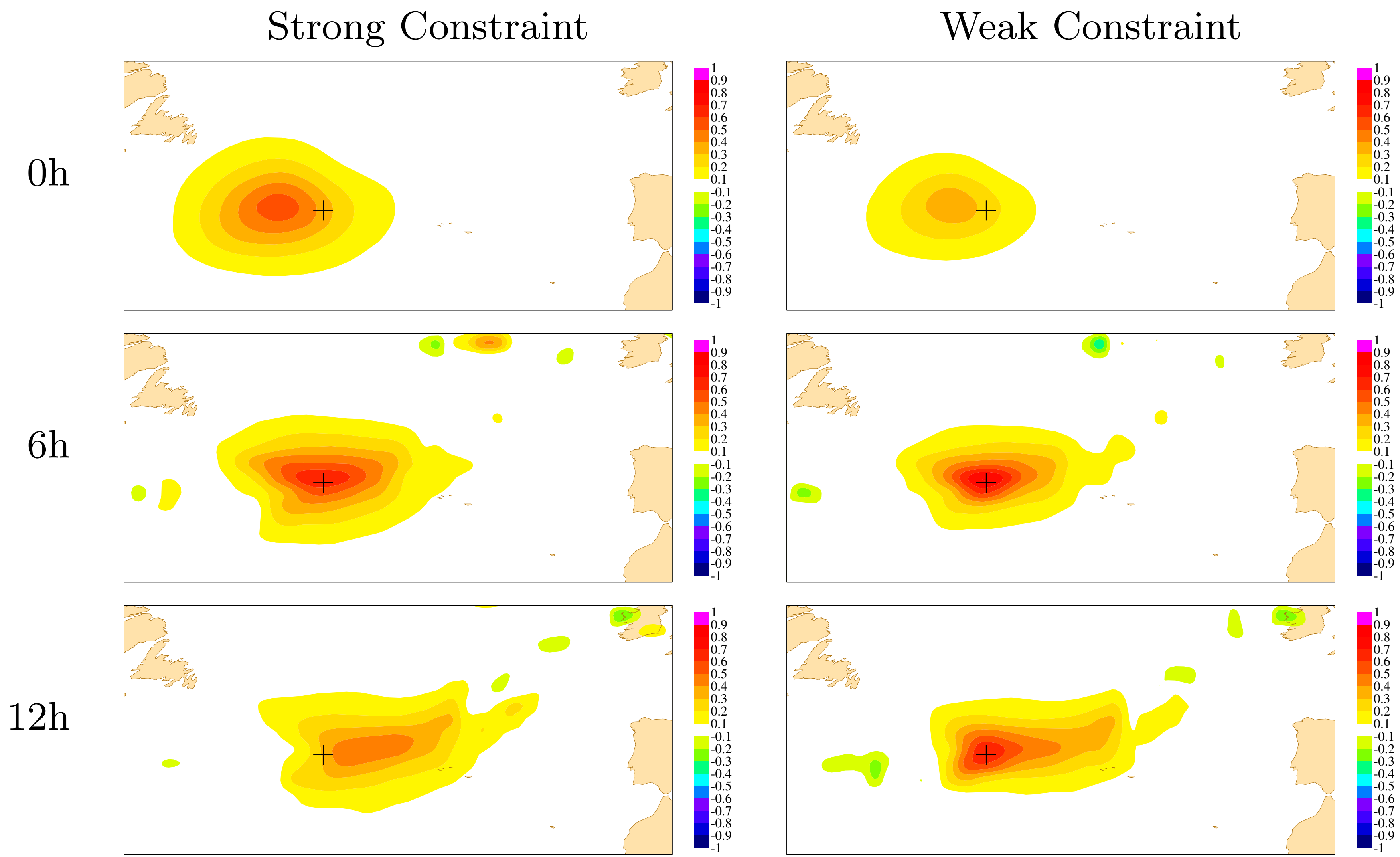
Aircraft Temperature Bias



Observations are biased.

Lars Isaksen

12 Observations Experiment



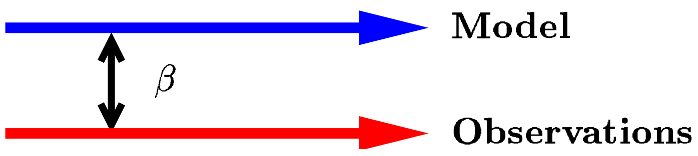
Systematic model error captures stationary errors.

Model Bias

- Model error β can be defined as: $\beta_i = x_i - \mathcal{M}_i(x_0)$
- Another control variable is $\chi' = (x_0, (\beta_i)_{i=1,\ldots,n})$,
- Approximation: $\beta_i = \beta$.
- The cost function is:

$$J_o(x) = \frac{1}{2}[\mathcal{H}(\mathcal{M}(x_0) + \beta) - y]^T R^{-1} [\mathcal{H}(\mathcal{M}(x_0) + \beta) - y]$$

- β is a 3D atmospheric state,
- The state of the model is not perturbed,
- β sees global model − observations bias,
- Does not correct for bias of one subset of observations against another subset of observations.



Model Bias Experiments

- Fits observations and background better,
- Observation statistics show some improvements and some degradations,
- Might absorb some of the signal,
- Added to forecast at post-processing stage,
- Scores slightly negative (needs proper verification),
- Q should have large scale correlations?
- Generate Q from prescribed error correlations length scales.

Summary

- Weak constraints 4D-Var is technically implemented in the IFS,
- Fits the data more uniformly over the assimilation window,
- Capture some model errors (winter stratosphere),
- Captures some observation bias (ascending/descending aircrafts),
- Sees systematic differences between model and observations:
  - Distinction between model bias and observation bias?
  - Interactions with variational observation bias correction?
- Two tools to define the errors we wish to capture:
  - Model error covariance matrix (scales, ...),
  - Model for model error (constant 3D state).

Future Work

- Compare various formulations of weak constraint 4D-Var (η, β, x),
- What type of model error (random vs. systematic vs. bias)?
- Determine appropriate model error covariance matrix,
- Interactions between model error and observation bias,
- Operational analysis and Reanalysis.